International Journal of Advanced Research in Biological Sciences
ISSN : 2348-8069
www.ijarbs.com

Research Article

Metagenomic analysis of Lonar soda lake samples

Thirumala Mothe* and Vishnuvardhan Reddy Sultanpuram

Department of Biochemistry, Mahatma Gandhi University, Anneparthy, Yellareddygudem (PO), Nalgonda 508254, India.
*Corresponding author: *thirumala_21@yahoo.com*

Abstract

Metagenomics is a new research area developed over the past decade, useful to identify potential enzymes and other compounds from nonculturable microbes with commercial interest. In the present work, Lonar crater samples were used to isolate Metagenomic DNA, which was used for shortgun whole metagenome sequencing. Shotgun sequencing technique was used to generate soil metagenomic profile. Raw data was uploaded to MG-RAST server for annotation, QC analysis and rRNA and gene prediction. The analysis showed that the metagenome sample contained a major annotated protein (39.2% with known function) and 1.3% of reads had similarity to ribosomal RNA genes. The metagenomic sequences belonged mostly to Bacteria (97.7%), with very less percentage of Archaea, Eukaryota and others.

Keywords: Metagenomics, Lonar crater, shortgun metagenome sequencing

Introduction

The majority of the microorganisms are unculturable with traditional techniques [1], So, obtaining novel genes from those unculturable environmental microorganisms became necessary [2]. It is now widely accepted that the soil metagenome is an important source of many useful biocatalysts (3). Recent studies have shown that one gram of soil may contain several thousand different species of microorganisms (4, 5, 6). However, 99% of the bacteria in soil cannot be cultured with conventional methods, leaving a large fraction of the soil microbial population unavailable for use (7). Metagenomics from various environments has been studied including gut microbiota, aquatic ecosystems, mines, agricultural and forest soils (8, 9, 10, 11, 12, 13, 14). These studies not only given the idea about the microbial ecology but also provided information about novel genes that could have applications in biotech industry (9, 10, 12, 15).

Metagenome isolation, preparation of clonal library and its screening is one method and direct whole metagenome shotgun sequencing of the isolated metagenome is another important method by which the genes present in the environmental samples are revealed. The latter method provides information both on variety of organisms present and the metabolic processes possible in the community (16). This could be helpful in understanding the microbial community, particularly when multiple samples are compared to each other. (17)

Previous attempts at DNA based analysis of soils used DNA fingerprinting techniques which evaluate fragment length variation such as terminal restriction fragment length polymorphism (TRFLP) (18), (19), denaturing gradient gel electrophoresis (DGGE) (20), amplified ribosomal DNA restriction analysis (ARDRA) (21) and length heterogeneity-polymerase chain reaction (LH-PCR) . All these methods have potential for use in forensic comparisons, however a lack of reproducibility and the potential for false inclusions has restricted their implementation in a forensic setting. Development of new platforms for

high-throughput DNA sequencing (HTS) has made it more affordable and led to the significant growth of HTS-based studies (22, 23, 24). Gene-targeted, or locus-specific, sequencing which typically targets the 16S rRNA gene is used for characterization of the taxonomic composition and diversity of microbial communities (25), (26). Shotgun sequencing is primarily a method for studying the functional structure of the communities which aims to examine the entire genetic assemblage and, being amplification-independent, relies on variation and commonality of the collective genomes found in a given environmental sample (27, 28).

Shotgun metagenomics is also capable of sequencing nearly complete microbial genomes directly from the environment (12). Because the collection of DNA from an environment is largely uncontrolled, the most abundant organisms in an environmental sample are most highly represented in the resulting sequence data. To achieve the high coverage needed to fully resolve the genomes of under-represented community members, large samples, often prohibitively so, are needed. On the other hand, the random nature of shotgun sequencing ensures that many of these organisms, which would otherwise go unnoticed using traditional culturing techniques, will be represented by at least some small sequence segments (12). Standard metagenomic approaches such as, shotgun and WGS are widely accepted as the most comprehensive sources of data for studying complex microbial communities. In this report, lonar soil samples were analysed using short gun metagenome sequencing method.

Materials and Methods

Sample collection

Samples were collected from the vicinity of the impact crater, Lonar lake, Buldhana, India (Latitude and Longitude – 19°58′36″N 76°30′30″E) in August 2013. The samples were collected in large quantities at various points to include the total biodiversity of the lake. The samples collected include, soil sediments and partially degraded wood chips collected from surrounding the lake. All the samples from different locations were randomly mixed together and a portion of the sample was used for metagenomic DNA isolation.

Metagenomic DNA Preparation

Metagenomic DNA was extracted from the samples collected from Lonar lake following the protocol from Zhou et al 1996 and as modified by (9). According to this protocol, 5g of sample was suspended in 10ml CTAB/proteinase K extraction buffer and incubate at 37˚C for 30 min. Then SDS to 2% final concentration was added and mixed gently before incubating the sample for 2 h at 60˚C, which was done with occasionally gentle inverting. Later, the sample was centrifuged at low speed (10 min at 6000g). Supernatant was transferred to a new tube and the sample extraction was repeated again with 5ml extraction buffer plus SDS. Carefully an equal volume of chloroform:isoamyl alcohol (24:1) was added to the pooled supernatant and was rotated gently for 10 min at ~15 rpm. Then the sample was centrifuged at low speed for 10min at 6000g. After the transfer of the aqueous (upper) layer to a new tube, 0.6 vol. isopropanol was added. Then the sample was mixed gently for 5 min and kept at room temperature (28℃) for 20 min. The sample was then centrifuged for 10 min at 12000g to pellet DNA and the pellet was washed with 70% ethanol before air drying it. At last the DNA was resuspend in 500µl of TE buffer and pipetted gently with wide-bore pipette and was incubated overnight at 4˚C. The DNA was tested for quality by using 1% agarose gel electrophoresis and Nanodrop (Thermo 5000) was used for estimating the quantity of the extracted DNA.

Processing, library construction and Illumina sequencing of the metagenomic DNA

3 µg of genomic DNA was sonicated with following conditions, 30s ON and 30s OFF at high intensity for 68 minutes to fragment DNA into size ranging between 200 to 500 bp (Bioruptor). The size distribution was checked by running an aliquot of the sample on Agilent BA-HS Chip. The resulting fragmented DNA was cleaned up using Agencourt AMPure XP SPRI beads (Beckman Coulter). Library preparation was performed following the NEXTFlex DNA library protocol outlined in “NEXTFlex DNA sample preparation guide (Cat # 5140-02)”. DNA was subjected to a series of enzymatic reactions that repair frayed ends, phosphorylate the fragments, and add a single nucleotide A overhang and ligate adaptors (NEXTFlex DNA Sequencing kit). Sample cleanup was done using AMPure SPRI beads. After ligation,

~300 – 600 bp fragment was size selected on a 2% agarose gel and cleaned using MinElute column (QIAGEN). 16cycles of PCR amplification of adaptor ligated fragments was done and cleaned up using AMPure SPRI beads. The prepared libraries were quantified using Qubit fluorometer and validated for quality by running an aliquot on High Sensitivity Bioanalyzer Chip (Agilent). The QC qualified metagenomic DNA library sample was sequenced using the Illumina Miseq platform. The above sequencing was performed at Genotypic Technologies Pvt. Ltd.

Data submission, QC and data annotation

The raw 7.10 GB raw data obtained through the Illumina Miseq platform sequencing was uploaded to MG – RAST (Metagenome Rapid Annotation using Subsystem Technology) server at http://metagenomics.anl.gov/ (30). The MG-RAST server is an automated analysis platform for metagenomes providing quantitative insights into microbial populations based on sequence data. The server primarily provides upload, quality control, automated annotation and analysis for prokaryotic metagenomic shotgun samples.

Results and Discussion

Using Illumina technique apporixmately 7.1 GB of data was generated in two runs with an average length of 179 bp. These metagenome sequences contains 3,539,070 high quality reads. The data was submitted to MG-RAST server for annotation (30)and analysis with the Accession ID 4560117.3.

Approximately 44,761 of the total 16S RNA gene encoding sequences, 3159 sequences could be identified upto species level. The proteobacteria 45.8%, Bacteroidetes 20.8%, Firmicutes 14.7% were the major phyla represented in the sample, remaining minor percentage were Actinobacteria, Aquificae , Ascomycota, Chlorobi, Chloroflexi, Chordata, Cyanobacteria, Deferribacteres, Deinococcus-Thermus, Euryarchaeota, Fusobacteria and others. The rank abundance curve is a tool for visually representing taxonomic richness and evenness which shows the phylum abundances ordered from the most abundant to least abundant (Figure 1). Numerous ecological studies show that soil microbial communities differ between land uses and vegetation types (31, 32, 33, 34, 35). The data generated by shotgun sequencing are commonly shifted towards describing the most abundant taxa leaving the contribution of rare microorganisms undervalued for comparative analysis (36).There is currently no agreed evaluation approach leading to an accurate picture of the soil metagenome structure as the true soil microbial community composition (37).

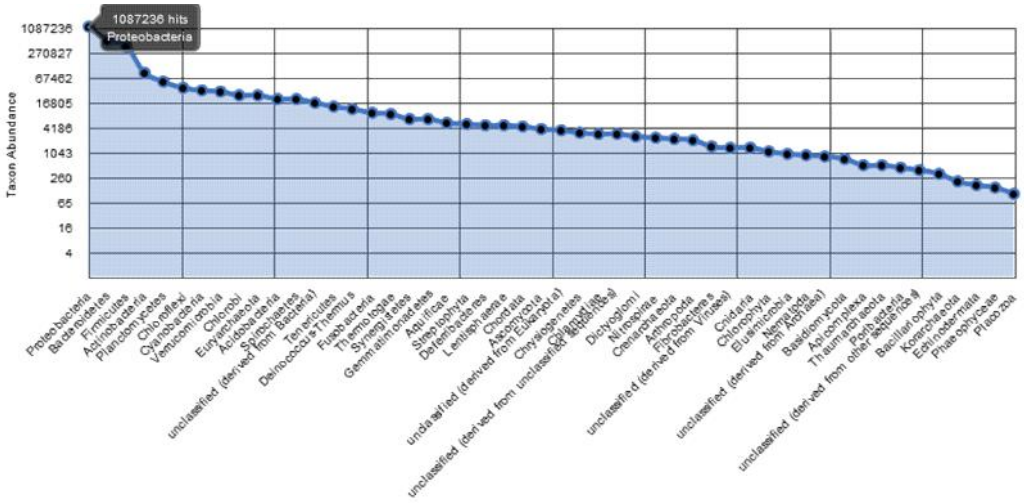

Figure 1: The rank abundance curve. The above plot shows the phylum abundances ordered from the most abundant to least abundant. The y-axis plots the abundances of annotations in each phylum on a log scale. Only the top 50 most abundant are shown here.

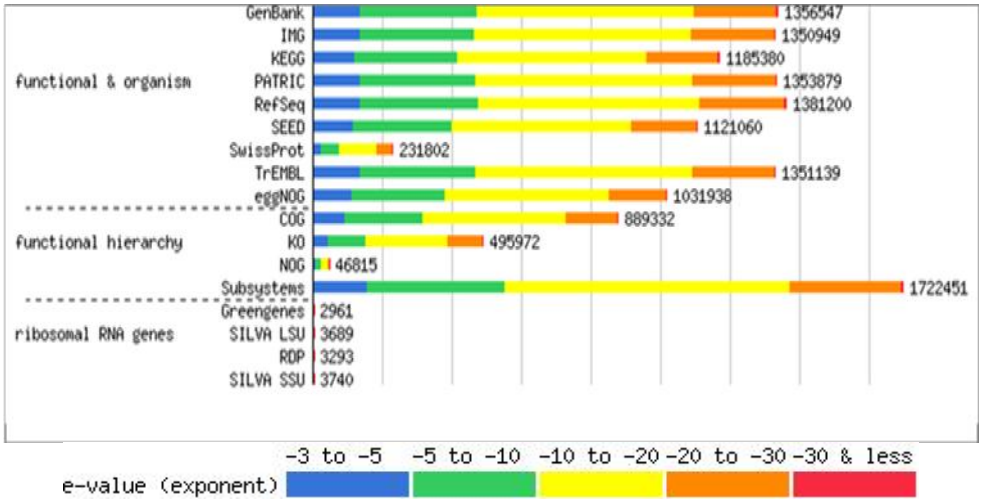

Figure 2: The graph displays the number of features in this dataset that were annotated by the different databases These include protein databases, protein databases with functional hierarchy information, and ribosomal RNA databases. 1,121,329 (39.2%) of the predicted protein features could be annotated with similarity to a protein of known function. 884,139 (78.8%) of these annotated features could be placed in a functional hierarchy. 44,761 (1.3%) of reads had similarity to ribosomal RNA genes.The bars representing annotated reads are colored by e-value range. The MG-RAST default annotation parameters such as maximum E-value $<1\times10^{-5}$, minimum length of alignment of 15 bp, and minimum sequence identity of 60%, were used to identify the best database matches.

Figure 2 shows that 1,121,329 (39.2%) of the predicted protein features could be annotated with similarity to a protein of known function. 884,139 (78.8%) of these annotated features could be placed in a functional hierarchy. 44,761 (1.3%) of reads had similarity to ribosomal RNA genes. The graph explains the number of features in the dataset that were annotated by the different databases such as, Genbank, IMG, KEGG, PATRIC, RefSeq, SEED, SwissProt,TrEMBL, eggNOG, COG, KO, NOG, Subsystems, Greengenes, SILVA LSU, RDP and SILVA SSU. These include protein databases, protein databases with functional hierarchy information which reveal that a majority of reads are useful for metabolism, (which involves so many enzymes like cellulases, lipases, proteases etc.) and ribosomal RNA databases. Different databases have different numbers of hits, but can also have different types of annotation data. There are 15,945,780 sequences in the M5NR protein database and 309,342 sequences in the M5RNA ribosomal database. The M5NR protein database contains all the unique sequences from the mentioned protein databases and the M5RNA ribosomal database contains all the unique sequences from the mentioned ribosomal RNA databases. Low levels of taxonomic or functional classification show less overlap between samples and are therefore also used frequently for metagenomic profile discrimination (38, 39). The results of the metagenomic dataset comparison in the current study are presented at all MG-RAST taxonomic and metabolic levels of hierarchy.

Conclusion

MT and SVR thank Department of Science & Technology, Government of India, for providing FTSYS project grants (SB/FT/LS -320/2012 and SB/FT/LS -115/2012).

References

1. Rappe MS, Giovannoni SJ (2003) The uncultured microbial majority. *Annu Rev Microbiol* 57:369–394
2. Daniel R (2004) The soil metagenome—a rich resource for the discovery of novel natural products. *Curr Opin Biotechnol* 15:199–204
3. Kowalchuk GA, Speksnijder AG, Zhang K, Goodman RM, van Veen JA (2007). Finding the needles in the metagenome haystack. *Microb Ecol* 53:475-485
4. Torsvik V, Goksøyr J, Daae FL (1990) High diversity in DNA of soil bacteria. *Appl Environ Microbiol* 56: 782-787
5. Chatzinotas A, Sandaa RA, Schönhuber W, Amann R , Daae FL, Torsvik V, Zeyer J, Hahn D (1998) Analysis of broad-scale differences in microbial community composition of two pristine forest soils. *Syst Appl Microbiol* 21: 579-587

6. Sessitsch A, Hackl E, Wenzl P, Kilian A, Kostic T, Stralis-Pavese N, Sandjong BT, Bodrossy L (2006) Diagnostic microbial microarrays in soil ecology. *New Phytol* 171: 719-735

7. Xu J (2006) Microbial ecology in the age of genomics and metagenomics: concepts, tools, and recent advances. *Mol Ecol* 15: 1713-1731

8. Kim U-J, Birren BB, Slepak T, Mancino V, Boysen C, Kang H-L, Simon MI, Shizuya H (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* 34: 213–218

9. Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR, Loiacono KA, Lynch BA, MacNeil IA , Tiong C Minor CL, Gilman M , Osburne MS, Clardy J , Handelsman J , Goodman RM (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* 66: 2541-2547

10. Wang GY, Graziani E, Waters B , Pan W, Li X, McDermott J, Meurer G , Saxena G, Andersen RJ, Davies J (2000) Novel natural products from soil DNA libraries in a streptomycete host. *Org Lett* 2: 2401–2404

11. Lee SW, Won K, Lim HK , Kim JC, Choi GJ, Cho KY (2004) Screening for novel lipolytic enzymes from uncultured soil microorganisms. *Appl Microbiol Biotechnol* 65: 720-726

12. Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM, Solovyev VV, Rubin EM, Rokhsar DS, Banfield JF (2004) "Insights into community structure and metabolism by reconstruction of microbial genomes from the environment". *Nature* 428 (6978): 37–43

13. Craig JW, Chang FY, Brady SF (2009) Natural products from environmental DNA hosted in *Ralstonia metallidurans ACS Chem Biol* 4: 23-28

14. Qin, J., Li R., Raes J., et al. (2010) A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464: 59-65

15. Nacke H, Will C, Herzog S, Nowka B, Engelhaupt M, Daniel R (2011) Identification of novel lipolytic genes and gene families by screening of metagenomic libraries derived from soil
samples of the German Biodiversity Exploratories. *FEMS Microbiol Ecol*, doi: 78:188-201

16. Nicola Segata, Daniela Boernigen, Timothy L Tickle, Xochitl C Morgan, Wendy S Garrett, Curtis Huttenhower (2013) "Computational meta'omics for microbial community studies". *Mol Syst Biol* 9 (666)

17. Allen EE, Banfield JF (2005) "Community genomics in microbial ecology and evolution". *Nature Rev Microbiol* 3 (6): 489–498

18. Macdonald CA, Ang R, Cordiner SJ, Horswell J (2011) Discrimination of soils at regional and local levels using bacterial and fungal T-RFLP profiling. *J Forensic Sci* 56: 61–69

19. Lenz EJ, Foran DR (2010) Bacterial profiling of soil using genus-specific markers and multidimensional scaling. *J Forensic Sci* 55: 1437–1442

20. Pasternak Z, Al-Ashhab A, Gatica J (2012) Optimization of molecular methods and statistical procedures for forensic fingerprinting of microbial soil communities. *Int Res J Microbiol* 3: 363–372

21. Concheri G, Bertoldi D, Polone E, Otto S, Larcher R et al (2011) Chemical elemental distribution and soil DNA fingerprints provide the critical evidence in murder case investigation. *PloS One* 6: e20222

22. Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Mol Ecol* 21: 1794–1805

23. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE et al (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30: 434–439

24. Logares R, Haverkamp THA, Kumar S, Lanzén A, Nederbragt AJ et al (2012) Environmental microbiology through the lens of high-throughput DNA sequencing: synopsis of current platforms and bioinformatics approaches. *J Microbiol Methods* 91: 106–113

25. Fierer N, Lauber CL, Ramirez KS, Zaneveld J, Bradford M A et al (2012) Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J* 6: 1007–1017

26. Suenaga H (2012) Targeted metagenomics: a high-resolution metagenomics approach for specific gene clusters in complex microbial communities. *Environ Microbiol* 14: 13–22

27. Delmont TO, Prestat E, Keegan KP, Faubladier M, Robe P et al (2012) Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J* 6: 1677–1687

28. Prakash T, Taylor TD (2012) Functional assignment of metagenomic data: challenges and applications. *Brief Bioinform* 13: 711–727

29. Zhou J, Bruns MA, Tiedje JM (1996) DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 62:316-322
30. Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM et al (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform* 9: 386
31. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates ST et al (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci USA* 109: 21390–21395
32. Xu Z, Hansen MA, Hansen LH, Jacquiod S, Sørensen SJ (2014) Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PloS One 9*: e93445.
33. Lauber CL, Ramirez KS, Aanderud Z, Lennon J, Fierer N (2013) Temporal variability in soil microbial communities across land-use types. *ISME J* 7: 1641–1650
34. Shange R, Haugabrooks E, Ankumah R, Ibekwe A, Smith R et al (2013) Assessing the Diversity and Composition of Bacterial Communities across a Wetland, Transition, Upland Gradient in Macon County Alabama. *Diversity* 5: 461–478
35. Uroz S, Ioannidis P, Lengelle J, Cébron A, Morin E et al (2013) Functional assays and metagenomic analyses reveals differences between the microbial communities inhabiting the soil horizons of a Norway spruce plantation. *PloS One* 8: e55929
36. Zarraonaindia I, Smith DP, Gilbert JA (2013) Beyond the genome: community-level analysis of the microbial world. *Biol Philos* 28: 261–282
37. Delmont TO, Simonet P, Vogel TM (2012) Describing microbial communities and performing global comparisons in the 'omic era. *ISME J* 6: 1625–1628
38. Jeffries TC, Seymour JR, Gilbert JA, Dinsdale E A, Newton K et al (2011) Substrate type determines metagenomic profiles from diverse chemical habitats. *PloS One 6*: e25173.
39. Håvelsrud OE, Haverkamp THA, Kristensen T, Jakobsen KS, Rike AG (2012) Metagenomic and geochemical characterization of pockmarked sediments overlaying the Troll petroleum reservoir in the North Sea. *BMC Microbiol* 12: 203